THE AUSTRALIAN NATIONAL UNIVERSITY

TR-CS-11-02

Acted Facial Expressions In The Wild Database

Abhinav Dhall, Roland Goecke, Simon Lucey, Tom Gedeon

September 2011

ANU Computer Science Technical Report Series

This technical report series is published by the Research School of Computer Science, College of Engineering and Computer Science, The Australian National University. Previously published by the School of Computer Science, and prior to 2009 this series was published as the Joint Computer Science Technical Report Series by the Department of Computer Science, FEIT, and the Computer Sciences Laboratory, RSISE, The Australian National University.

Please direct correspondence regarding this series to:

Technical Reports
Research School of Computer Science
College of Engineering and Computer Science
The Australian National University
Canberra ACT 0200
Australia

or send email to:

`Technical-DOT-Reports-AT-cs-DOT-anu.edu.au`

A list of technical reports, including some abstracts and copies of some full reports may be found at:

`http://cs.anu.edu.au/techreports/`

Recent reports in this series:

TR-CS-11-01 Muhammad Atif and Peter Strazdins. *Adaptive Resource Remapping In Virtualized Environments - Framework*. May 2011.

TR-CS-10-01 Duncan Stevenson, Garry Warne, Mai Eames, Maurice Hennessy, Dan Penny, Jim Wilkinson, and Simon Pase. *Delivery of Postgraduate Medical Training to Developing Countries - A Pilot Study*. September 2010.

TR-CS-09-02 Daniel Frampton, Stephen M Blackburn, Luke N Quinane, and John N Zigman. *Efficient Concurrent Mark-Sweep Cycle Collection*. October 2009.

TR-CS-09-01 Peter Christen, Ross Gayler, and David Hawking. *Similarity-Aware Indexing for Real-Time Entity Resolution*. August 2009.

TR-CS-08-03 Jie Cai, Alistair P. Rendell, Peter E. Strazdins, and H'sien Jin Wong. *Predicting Performance of Intel Cluster OpenMP with Code Analysis Method*. November 2008.

TR-CS-08-02 Paul Thomas. *Implementation of PIS*. June 2008.

Acted Facial Expressions In The Wild Database

Abhinav Dhall[1], Roland Goecke[2,1], Simon Lucey[3] and Tom Gedeon[1]

[1] Research School of Computer Science, Australian National University, Canberra, Australia
[2] Faculty of Information Sciences and Engineering, University of Canberra, Australia
[3] Commonwealth Scientific and Industrial Research Organisation (CSIRO), Australia
Email: abhinav.dhall@anu.edu.au, roland.goecke@ieee.org, simon.lucey@csiro.au, tom.gedeon@anu.edu.au

Abstract. Quality data recorded in varied realistic environments is vital for effective human face related research. Currently available datasets for human facial expression analysis have been generated in highly controlled lab environments. We present a new dynamic 2D facial expressions database based on movies capturing diverse scenarios. A new XML schema based approach has been developed for the database collection and distribution tools. Our database captures varied facial expressions, natural head pose movements, occlusions, subjects from various races, gender, diverse ages and multiple subjects in a scene. The subjects have been labelled with information such as name, age of character, age of actor, gender, pose and individual facial expressions. In total, we used thirty-seven movie DVDs. The database consists of a detailed XML schema which contains information about the clips and experiment protocols.

1 Introduction

Realistic face data plays a vital role in the research advancement of facial expression analysis systems. Much progress has been made in the fields of face recognition and human activity recognition in the past years due to the availability of realistic databases as well as robust representation and classification techniques. Inspired by them, we present a labelled temporal facial expression database from movies. Human facial expression databases till now have been captured in controlled 'lab' environments. Our database constitutes information on clips of both single and multiple subjects interacting with each other. To overcome the tedious and error-prone process of manual data collection, we use a semi-automatic method based on searching the Subtitles for Deaf and Hearing impaired (SDH) as well as Closed Caption (CC). The database covers unconstrained facial expressions, varied head poses and movements, vast age range, occlusions, varied focus, multiple people in the same scene and close-to-real world illumination. To the best of our knowledge, all existing facial expression databases till date have a single subject in focus whereas our contains single and multiple subjects. The range of age of subjects in the clips is large from 1 - 70 years. The information about the clips has been stored in an extensible XML schema and the subjects in the clips have been annotated with attributes such as *Name*, *Age of Actor*, *Age of Character*, *Pose*, *Gender*, *Expression of Person* and the overall *Clip Expression*.

Fig. 1. Sample frames from our database AFEW. Picture 1 and 2 are from the movie 'About a Boy'. Picture 3 and 4 are from the movie 'The Hangover' and Picture 5 is from the movie 'It's Complicated'. Please note the diverse nature and scenarios of the snapshots.

We have named our database *Acted Facial Expressions in the Wild* similar to the spirit of the Labeled Faces in the Wild (LFW) database [1]. It contains 957 videos labelled with six basic expressions *Angry*, *Happy*, *Disgust*, *Fear*, *Sad*, *Surprise* and the *Neutral* expression. We also wanted to capture the information on how facial expressions evolved in subjects with age. Therefore we have chosen sets of movies featuring the same actors. For example, the *Harry Potter* series forms a good platform to analyse how facial expressions of subjects evolve with age. We used thirty-seven movies from a diverse range of movie genres so as to cover as much varied expressions and natural environments as possible. Figure 1 contains some screenshots from the database.

AFEW is an acted facial expressions dataset in tough conditions. Given that there is a huge amount of video data on the web which is posed, it is worthwhile to investigate the problem of facial expressions analysis in tough conditions (unlike lab conditions). Ideally, we would like a dataset of facial expressions in challenging real-world environments. However, as anyone in the face analysis community will attest, such datasets are extremely difficult to obtain. As an important stepping stone on this path, we therefore propose a dataset of temporal facial expressions extracted from actors in movies (in environments similar to those found in the real world). For the transition of facial expression analysis approaches from labs to realistic environments, we need databases which can mimic the real world.

While movies are often shot in somewhat controlled environments, they provide close to real world environments that are much more realistic than current datasets that were recorded in lab environments. We are not claiming that the AFEW database is a spontaneous facial expression database. However, clearly, (good) method actors attempt mimicking real world human behaviour such that audiences get the illusion that they are spontaneous not posing in movies. The dataset in particular addresses the issue of temporal facial expressions in difficult conditions that are approximating real world conditions, which provides for a much more difficult test set than currently available datasets.

Automated human facial expression analysis is the study of computer based understanding of local muscle movements on human faces induced by internal emotional states, intentional, or social communications. Facial expression analysis has a variety of applications in Human Computer Interaction (HCI) and affective computing. It is a widely researched but still open problem. On the bases of the descriptor type, facial expression analysis methods can be divided into three categories: geometric based [2], [3], [4], appearance based [5], [6], [7] and combined [8]. Further, facial expression analysis methods can also be classified into image based [9], [10] and video based [11], [12]. A detailed survey on various facial expression analysis methods can be found in [13], [7]. The current state of the art facial expression analysis methods such as [7], [12], [14] and [15] have shown robust performance over the current databases. However all temporal databases are lab controlled and, hence, do not really represent real world environments. Training and testing on real world environment data sets is mandatory for the advancement of facial expression analysis. This is the main motivation behind the AFEW database. The database provides movie clips of facial expressions in real-world or close to real-world, situations and environments.

Database	Static /Dynamic	Lab Environment	Age Range	Profile View	Illum.	Occlusion	No. of Subjects	Audio	Searchable
AFEW	Dynamic	Real	1-70	Yes	CTN	Yes	220	Yes	Yes
Belfast [24]	Dynamic	TV & Lab	?	No	C	Yes	100	No	No
CK [17]	Both	Lab	18-50	No	C	No	97	No	No
CK+ [18]	Both	Lab	18-50	Yes	C	No	123	No	No
Facial Age [26]	Static	Real	0-69	Yes	CTN	Yes	82	No	No
F.TUM [25]	Both	Lab	?	No	C	No	18	No	No
GEMEP [27]	Dynamic	Lab	?	Yes	C	Yes	10	No	No
LFW [1]	Static	Real	?	Yes	CTN	Yes	5749	No	No
M-PIE [21]	Both	Lab	27.9	Yes	C	Yes	337	No	No
MMI [19]	Both	Lab	19-62	Yes	C	Yes	19	No	Yes
PIE [20]	Static	Lab	?	Yes	C	Yes	68	No	No
RU-FACS [23]	Dynamic	Lab	18-30	No	C	Yes	100	No	No
UT-Dallas [22]	Both	Lab	18-25	Yes	C	Yes	284	No	No

Table 1. Comparison of facial expression databases. Here, C = Controlled, CTN = Close to natural.

2 Related Databases

Facial expression analysis has been researched actively in the past few decade. A variety of human facial expression databases were introduced including both static and dynamic data and posed and spontaneous expressions. The Facial Action Coding System (FACS) [16] is an expression coding system developed by Paul Ekman, a behavioural scientist. It decomposes human faces movements into 46 component movements, which are referred to as Action Unit (AU). It has been used extensively in both automatic and manual facial expression analysis research.

The Cohn-Kanade database [17] is widely used facial expression database. This database formed a standard for facial expression recognition, it contains both static and dynamic data captured under lab conditions. It has 486 lab recorded sequences and 97 subjects. The CMU Pose, Illumination, and Expression (PIE) Database [20] is a collection of 40,000 facial images of 68 subjects covering various illuminations and poses recorded under controlled lab conditions. This database has been used extensively in the literature for facial expression analysis and face recognition. The Multi-PIE database [21] was developed to overcome the limitations of the PIE database and consists of 337 subjects, 15 view points and 19 different illumination conditions recorded under controlled lab conditions. Like PIE, it contains both static and temporal data. The MMI database [19] is a FACS coded, posed facial expression database containing static and dynamic data. It has a web driven update and download model but has limited subjects viz 19. This is the only database which has a search provision like AFEW.

All databases discussed above are posed facial expression analysis databases. The RU-FACS database [23] is a FACS coded temporal database where 100 subjects participated in a false opinion paradigm. This is a spontaneous facial expression database, but it is proprietary and unavailable to other researchers. All these databases though have been recorded in lab controlled restricted environments. The Belfast database [24] consists of a combination of studio recordings and TV programme grabs labelled with particular emotions. The number of TV clips in this database are few in number.

The Labeled Faces in the Wild database (LFW) [1] is a static face database created from face images found on the internet. It contains natural head poses, varied illumination, age, gender and occlusion. Similar to LFW is the Pubfig database [28], which contains 58797 images of 200 people collected from the internet. In [29], and [30], a face database was created by extracting faces every 10th frame from two movies using the Viola-Jones face detector [31]. Though these databases cover varied scenarios and real world environment, all of them are static in nature.

Laptev et al., introduced the Hollywood Human Actions (*HOHA*) dataset [32] and the Hollywood-2 [33] database. The databases have been extracted from Hollywood movies using time stamp information in the subtitles and the corresponding keywords in the movie scripts. The dataset has been automatically labelled for action classes such as 'AnswerPhone', 'GetOutCar' and scene information such as 'INT-BEDROOM', 'INT-CAR' etc. These are extracted using machine learning techniques for text classification described in [34]. Their technique of extracting video clips from movies have some similarities to our method, but the method of extraction, labelling and intended data distribution are different. We also use the time stamps from the subtitles but we extract the subtitles and use expression keywords for example '[CHEERS]', '[SHOUTS]' etc., in them unlike HOHA in which the movie scripts have been used. Also we provide a very detailed annotation about the subjects, which is vital information for affective computing and expression analysis. In Table 1, a detailed comparison of facial expression databases is presented.

3 Contributions of the database

The Acted Facial Expressions in the Wild database has the following novelty:

- The database is a dynamic temporal facial expressions data corpus consisting of close to real world environments unlike most other current facial expression databases recorded in a lab environment.
- The subjects in the database exhibit natural (including out-of-plane) head poses and movements, which are largely missing in other current temporal facial expression databases.
- The subjects belong to a wide range of ages from 1 to 70 years old, which makes it very generic in terms of age, unlike other facial expression databases. The database has a large number of clips depicting children and teenagers, which can be used to study facial expressions in children. Databases for facial age analysis usually have a similar large range of age but are static databases [26],[35]. Our dataset can also be used for both static and temporal facial age research.

- AFEW is currently the only facial expression database, which has multiple labelled subjects in the same frame. This enables an interesting study on the 'theme' expression of a scene with multiple subjects, which may or may not have the same expression at a given time.
- The database exhibits 'close to real' illumination conditions. The clips have various illumination scenarios such as indoor, night-time and outdoor natural illumination. While it is clear that movie studios use very controlled illumination conditions even in outdoor settings, we argue that these are closer to natural conditions than lab-controlled conditions and therefore valuable for facial expression research. The diverse nature of the illumination conditions in the dataset makes it useful for not just facial expression analysis but potentially also for face recognition, face alignment, age analysis and action recognition.
- The design of the database schema is based on XML. This enables further information about the data and its subjects to be added at any stage without changing the video clip. This means that detailed annotations with attributes about the subjects and the scene are possible. In the current version, the schema covers *Name*, *Age of Actor*, *Age of Character*, *Pose*, *Gender*, *Expression of Person* and the overall *Clip Expression*, which will be extended keeping the same video clips.
- The movies have been chosen covering a large set of actors. Many actors have appeared in multiple movies in the dataset, which will enable to research on how their expressions have evolved over the time, whether they differ for different genres, etc.

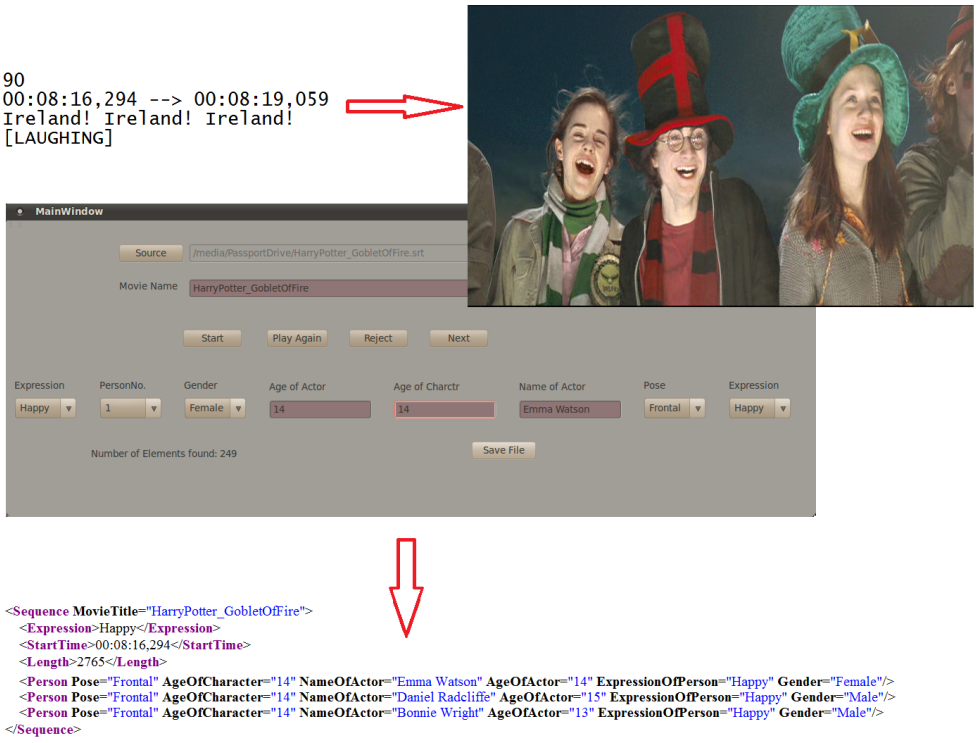
4 Database creation

A semi-automatic approach was followed during the creation of the database. We bought and analysed thirty-seven DVD movies. The whole system of the database creation was divided into two steps. First we parse the subtitles from the movies. The subtitles are searched for expression keywords. Then the time stamps of the subtitles containing keywords are used to play video clips. Then in the second step, a human observer annotates the clips with the information about the actors and the expressions. Figure 2 illustrates the database creation process.

4.1 Subtitle Parsing

Subtitles contain very useful information about the video. We used Subtitles for Deaf and Hearing impaired (SDH) and Closed Captions (CC) subtitles for database creation. The difference between SDH and CC is that, SDH subtitles have the same font as the translation subtitle; however CC are displayed as white text over black background. However both these types of subtitles contain information about the audio and non-audio context such as emotions, information about the actors and the scene for example '[CHEERING]', '[SHOUTS]', '[SURPRISED]' etc. We extracted the subtitles from the movies using a tool called VSRip[4]. For the movies where VSRip could not extract

[4] The tool used was VSRip from http://www.videohelp.com/tools/VSRip. It extracts .sub/.idx from DVD movies




```
<Sequence MovieTitle="HarryPotter_GobletOfFire">
  <Expression>Happy</Expression>
  <StartTime>00:08:16,294</StartTime>
  <Length>2765</Length>
  <Person Pose="Frontal" AgeOfCharacter="14" NameOfActor="Emma Watson" AgeOfActor="14" ExpressionOfPerson="Happy" Gender="Female"/>
  <Person Pose="Frontal" AgeOfCharacter="14" NameOfActor="Daniel Radcliffe" AgeOfActor="15" ExpressionOfPerson="Happy" Gender="Male"/>
  <Person Pose="Frontal" AgeOfCharacter="14" NameOfActor="Bonnie Wright" AgeOfActor="13" ExpressionOfPerson="Happy" Gender="Male"/>
</Sequence>
```

Fig. 2. The screenshot described the process of database formation. For example in the screenshot, when the subtitle contains the keyword ‘laughing’, the corresponding clip is played by the tool. The human labeller then annotates the subjects in the scene using the GUI tool. The resultant annotation is stored in the XML schema shown in the bottom part of the snapshot. Please note the structure of the information about a sequence containing multiple subjects. The image in the screenshot is from the movie ‘Harry Potter and The Goblet Of Fire’.

subtitles, we downloaded the SDH subtitles from the internet[5]. The extracted subtitle images were parsed using Optical Character Recognition (OCR) and converted into .srt subtitle format[6]. The .srt format contains the start time, end time and text content with millisecond accuracy.

4.2 Expression clip and subject labelling

A Qt based GUI tool was developed for data labelling. The tool performs a regular expression search with keywords[7] describing expressions and emotions on the subtitle file. This gives a list of subtitles with time stamps which contain information about some expression. The extracted subtitles containing expression related keywords were then played by the tool sequentially. The duration of each clip is equal to the time period of appearance of subtitle on the screen. The human observer then annotated the played video clips with information about the subjects[8] and expressions. In the case of video clips which have multiple actors, the sequence of labeling was based on two criteria. For actors appearing in the same frame, the ordering of annotation is left to right. When the actors appear at different time stamps, then it is the order of appearance. The labeling is then stored in the metadata schema. Also, the observer estimated the age of the character in most of the cases as the age of all characters in a particular movie was not available on the internet for all the characters.

4.3 XML based database metadata

Our database consists of video clips and metadata about the video clips. We developed an Extensible Markup Language (XML) based schema, which enables efficient data handling and updating. The human labellers annotated the subjects in the clips with detailed information. The schema elements are described as follows:

- *Expression* - This specifies the theme expression conveyed by the scene. The expressions were divided into six expression classes plus neutral. The default value is based on the search keyword found in the subtitle text, for example for smile and cheer is Happy. The human observer can change it based on their observation of the audio and scene of the clip.
- *StartTime* - This denotes the start timestamp of the clip in the movie DVD and is in the hh:mm:ss,zzz format.
- *Length* - It is the duration of the clip in milliseconds.
- *Person* - This contains various attributes describing the actor in the scene.

[5] The SDH subtitles were downloaded from www.subscene.com, www.mysubtitles.org and www.opensubtitles.org.

[6] We used Subtitle Edit available at www.nikse.dk/se, which can convert among various subtitle format.

[7] Keyword examples: [HAPPY], [SAD], [SURPRISED], [SHOUTS], [CRIES], [GROANS], [CHEERS] etc.

[8] The information about the actors was extracted from www.imdb.com.

- *Pose* - This denotes the pose of the actor, based on the human labeller's observation. In the current version we are manually classifying the head pose as frontal and non-frontal.
- *AgeOfCharacter* - This describes the age of the character based on human labeler's observation. In a few cases where the age of the character was available www.imdb.com, this information was used. Frequently this was only the case for lead actors.
- *NameOfActor* - This attribute contains the real name of the actor.
- *AgeOfActor* - This describes the real age of the actor. The information was extracted from www.imdb.com by the human labeller. In a very few cases the age information was missing for some actors, therefore, the observational values were used.
- *ExpressionOfPerson* - This denotes the expression category as labelled by the human observer. This may be different from the 'Expression' tag as there may be multiple people in the frame showing different expressions with respect to each other and the scene/theme.
- *Gender* - This attribute describes the gender of the actor, again entered by the human labeler.

This XML based metadata structure has two major advantages. First and foremost, it is easy to use and search using any standard programming language which supports XML. Secondly, the structure makes it simple to add new attributes about the video clips such as pose of the person in degrees and scene information, in the future. All this can be added keeping the existing data and making sure that the already existing tools which use it can take advantage of this information with minor changes. Currently, the database metadata indexes 957 video clips. Details of the database are in the Table 2. The details on how to get the database and its experiment protocols will be made available on:

http://cs.anu.edu.au/few

4.4 Movies in the database

The movies used in the database are: 21, About a boy, American History X, Black Swan, Did You Hear About The Morgans?, Dumb And Dumberer: When Harry Met Lloyd,

Attribute	**Description**
Length of sequences	300-5400 ms
No. of sequences	957
No. of annotators	1
Expression classes	Angry, Disgust, Fear, Happy, Neutral, Sad and Surprise
Total No. of expressions (some seq. have multiple subjects)	1259
Video format	AVI

Table 2. Attributes of AFEW database.

Protocol	**SFEW**	**AFEW**
	Train-Test sets have the:	
SPS	same single subject	same single subject
PPI	seen & unseen subjects	seen & unseen subjects
SPI	unseen subjects (Part of BEFIT workshop [36])	unseen subjects

Table 3. Different training and testing protocol scenarios for SFEW and AFEW. *SPS* - Strictly Person Specific, *PPI* - Partial Person Independent, *SPI* - Strictly Person Independent.

Four weddings and a Funeral, Frost/Nixon, Harry Potter and the Philosopher's Stone, Harry Potter and the Chamber of Secrets, Harry Potter and the Goblet of Fire, Harry Potter and the Half Blood Prince, Harry Potter and the Order Of Phoenix, Harry Potter and the Prisoners Of Azkaban, It's Complicated, I Think I Love My Wife, Little Manhattan, Notting Hill, One Flew Over The Cuckoo's Nest, Pretty In Pink, Pretty Woman, Remember Me, Runaway Bride, Saw 3D, Serendipity, Terms of Endearment, The Aviator, The Devil Wears Prada, The Hangover, The Informant!, The King's Speech, The Social Network, The Terminal, The Town, Valentine Day, Unstoppable, You've Got Mail.

5 Future Work

The structure of the metadata in AFEW makes it simpler to add new attributes about the subjects and scenes. The future work for the database is as follows:

- Creating a static dataset (SFEW) [36] from AFEW. And defining training and testing protocols for the database as defined in the Table 3.
- Using the location (x,y) of the faces in the image in a scene can be helpful for accurate initialisation for feature extractors. We will add face location information using face localisation methods such as Viola-Jones [31] and Constrained Local Models (CLM). [37]. We will also provide dense landmark annotation of the faces using person dependent AAMs [38].
- Using a head pose estimator, we will provide accurate pose angles.
- Contextual information about the scene such as 'Park', 'Room' etc., will be added to the metadata for enabling the study of the effect of scene context on expressions.
- Humans exhibit expressions with various intensities. The expression intensities will be added along with the 'ExpressionOfPerson'. In addition, FACS [16] annotations can be added at the frame level.
- The database currently has many occlusion scenarios such as glasses, eye mask, beard, hand and hat (examples in Figure 1). Information about the presence of occlusion and its type will be added as an attribute to the metadata.

6 Conclusions

We have presented a new dynamic temporal facial expression database derived from movies. The database contains natural head poses and movements, close to real world

illumination, multiple subjects in same frame, a large age range, occlusions and searchable metadata. It is the first database with realistic temporal video scenario for facial expression analysis. It also covers toddler, child and teenager subjects, which are missing in the currently available temporal facial expression databases. It also contains clips with multiple subjects exhibiting similar or different expressions with respect to the scene and each other. The project website will be used to post protocols and baselines. We hope that the dataset will be useful for researchers and play a vital role in the advancement of facial expression analysis in tough conditions.

References

1. Gary B. Huang, Manu Ramesh, Tamara Berg, and Erik Learned-Miller. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical Report 07-49, University of Massachusetts, Amherst, 2007.
2. Akshay Akshay Asthana, Jason Saragih, Micheal Wagner, and Roland Goecke. Evaluating AAM Fitting Methods for Facial Expression Recognition. In *Proceedings of the IEEE International Conference on Affective Computing and Intelligent Interaction*, ACII'09, pages 598–605, 2009.
3. Zhihong Zeng, M. Pantic, G.I. Roisman, and T.S. Huang. A Survey of Affect Recognition Methods: Audio, Visual, and Spontaneous Expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 39–58, 2009.
4. Abhinav Dhall, Akshay Asthana, and Roland Goecke. Facial expression based automatic album creation. In *Proceedings of the 17th international conference on Neural information processing: models and applications - Volume Part II*, ICONIP'10, pages 485–492, 2010.
5. Abhinav Dhall, Asthana Asthana, Roland Goecke, and Tom Gedeon. Emotion recognition using PHOG and LPQ features. In *Proceedings of the Ninth IEEE International Conference on Automatic Face Gesture Recognition and Workshops (FG'2011), Facial Expression Recognition and Analysis Challenge Workshop (FERA)*, pages 878–883, 2011.
6. Liefei Xu and Philippos Mordohai. Automatic facial expression recognition using bags of motion words. In *Proceedings of the British Machine Vision Conference*, BMVC'10, pages 1–13, 2010.
7. Jacob Whitehill, Gwen Littlewort, Ian R. Fasel, Marian Stewart Bartlett, and Javier R. Movellan. Toward practical smile detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 2106–2111, 2009.
8. Simon Lucey, Iain Matthews, Changbo Hu, Zara Ambadar, Fernando de la Torre, and Jeffrey Cohn. AAM Derived Face Representations for Robust Facial Action Recognition. In *Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition*, FG'2006, pages 155–162, 2006.
9. Maya Pantic and L.J.M. Rothkrantz. Facial action recognition for facial expression analysis from static face images. *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, pages 1449–1461, 2004.
10. Zisheng LI, Jun ichi IMAI, and Masahide Kaneko. Facial-component-based bag of words and PHOG descriptor for facial expression recognition. In *Proceedings of the 2009 IEEE international conference on Systems, Man and Cybernetics*, SMC'09, pages 1353–1358, 2009.
11. Michel Valstar and Maja Pantic. Fully automatic facial action unit detection and temporal analysis. In *Proceeding of IEEE International Conference on Computer Vision and Pattern Recognition Workshop*, CVPR-W'06, pages 149–149, 2006.

12. Feng Zhou, Fernando de la Torre, and Jeffrey F. Cohn. Unsupervised Discovery of Facial Events. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, CVPR'10, pages 2574–2581, 2010.
13. Beat Fasel and Juergen Luettin. Automatic Facial Expression Analysis: A Survey. *Pattern Recognition*, 36:259–275, 2003.
14. Guoying Zhao and Matti Pietikainen. Dynamic texture recognition using local binary patterns with an application to facial expressions. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 2007.
15. Peng Yang, Qingshan Liu, and Dimitris N. Metaxas. Exploring facial expressions with compositional features. In *Proceedings of the International Conference on Computer Vision and Pattern Recognition*, CVPR'10, pages 2638–2644, 2010.
16. P. Ekman and W.V. Friesen. The facial action coding system: A technique for the measurement of facial movement. In *Consulting Psychologists*, 1978.
17. Takeo Kanade, Jeffrey F. Cohn, and Yingli Tian. Comprehensive database for facial expression analysis. In *Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition*, FG'00, pages 46–53, 2000.
18. Patrick Lucey, Jeffrey F. Cohn, Takeo Kanade, Jason Saragih, Zara Ambadar, and Iain Matthews. The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression. In *CVPR4HB10*, 2010.
19. Maja Pantic, Michel François Valstar, Ron Rademaker, and Ludo Maat. Web-based database for facial expression analysis. In *Proceedings of the IEEE International Conference on Multimedia and Expo*, ICME'05, pages 317–321, 2005.
20. Terence Sim, Simon Baker, and Maan Bsat. The CMU Pose, Illumination, and Expression Database. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(12):1615–1618, December 2003.
21. Ralph Gross, Iain Matthews, Jeffrey F. Cohn, Takeo Kanade, and Simon Baker. Multi-PIE. In *Proceedings of the Eighth IEEE International Conference on Automatic Face and Gesture Recognition*, FG'2008, pages 1–8, 2008.
22. Alice J. O'Toole, Joshua Harms, Sarah L. Snow, Dawn R. Hurst, Matthew R. Pappas, Janet H. Ayyad, and Herve Abdi. A video database of moving faces and people. *IEEE TPAMI*, 2005.
23. Marian Stewart Bartlett, Gwen Littlewort, Mark G. Frank, Claudia Lainscsek, Ian R. Fasel, and Javier R. Movellan. Automatic recognition of facial actions in spontaneous expressions. *Journal of Multimedia*, 1(6):22–35, 2006.
24. Ellen Douglas-Cowie, Roddy Cowie, and Marc Schrder. A new emotion database: Considerations, sources and scope. In *Speech Comm.*, 2000.
25. Frank Wallhoff. Facial expressions and emotion database, 2006. `http://www.mmk.ei.tum.de/ waf/fgnet/feedtum.html`.
26. "Face and Gesture Recognition Research Network.". Fg net aging database. http://www.fgnet.rsunit.com.
27. T. Bänziger and K.R. Scherer. Introducing the Geneva Multimodal Emotion Portrayal (GEMEP) Corpus. In K.R. Scherer, T. Bänziger, and E. Roesch, editors, *Blueprint for affective computing: A sourcebook*. Oxford, England: Oxford University Press, 2010.
28. Neeraj Kumar, Alexander C. Berg, Peter N. Belhumeur, and Shree K. Nayar. Attribute and Simile Classifiers for Face Verification. In *Proceedings of the IEEE International Conference of Computer Vision*, ICCV'09, 2009.
29. Ognjen Arandjelovic and Andrew Zisserman. Automatic Face Recognition for Film Character Retrieval in Feature-Length Films. In *Proceedings of International Conference on Computer Vision and Pattern Recognition*, CVPR'05, pages 860–867, 2005.
30. Ognjen Arandjelovic and Roberto Cipolla. Automatic cast listing in feature-length films with anisotropic manifold space. In *Proceedings of International Conference on Computer Vision and Pattern Recognition*, CVPR'06, pages 1513–1520, 2006.

31. Paul A. Viola and Michael J. Jones. Rapid object detection using a boosted cascade of simple features. In *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*, CVPR'01, pages 511–518, 2001.
32. Ivan Laptev, Marcin Marszałek, Cordelia Schmid, and Benjamin Rozenfeld. Learning realistic human actions from movies. In *Proceedings of International Conference on Computer Vision and Pattern Recognition*, CVPR'08, pages 1–8, 2008.
33. Marcin Marszalek, Ivan Laptev, and Cordelia Schmid. Actions in context. In *Proceedings of International Conference on Computer Vision and Pattern Recognition*, CVPR'09, pages 2929–2936, 2009.
34. Fabrizio Sebastiani. Machine learning in automated text categorization. *ACM Comput. Surv.*, 34(1):1–47, 2002.
35. Meredith Minear and Denise C Park. A lifespan database of adult facial stimuli. *Behav Res Methods Instrum Comput*, 36(4):630–3, 2004.
36. Abhinav Dhall, Roland Goecke, Simon Lucey, and Tom Gedeon. Static Facial Expression Analysis In Tough Conditions: Data, Evaluation Protocol And Benchmark. In *IEEE International Conference on Computer Vision workshop, BEFIT (Accepted for publication)*, BEFIT'11, 2011.
37. Jason M. Saragih, Simon Lucey, and Jeffrey Cohn. Face alignment through subspace constrained mean-shifts. In *Proceedings of the IEEE International Conference of Computer Vision,*, ICCV'09, 2009.
38. Jason Saragih and Roland Goecke. Learning AAM fitting through simulation. *Pattern Recognition*, 42(11):2628–2636, 2009.